

КИТАЙСКО-РУССКИЙ ПАРАЛЛЕЛЬНЫЙ ДИСКУРСИВНЫЙ КОРПУС: ВЫРАВНИВАНИЕ НА УРОВНЕ КЛАУЗЫ И СТАТИСТИЧЕСКИЙ АНАЛИЗ

Статья посвящена способам корпусного выравнивания и аннотирования параллельных текстов на уровне клаузы. Работа строится на основе создания китайско-русского дискурсивного параллельного корпуса. В качестве пилотного материала был выбран китайский официальный документ «Доклад о работе правительства КНР 2017» и его перевод на русский язык. Данная работа состоит из трех общих частей: 1) характеристика принципов выравнивания двух текстов; 2) установление принципов аннотирования выделяемых единиц и создание параллельного корпуса, в котором выравнивание осуществляется на уровне клаузы; 3) статистика по разным типам русских синтаксических аналогов китайских клауз и интерпретация полученной информации. В первой части рассматривается, как в параллельном корпусе производится деление текстов на клаузы: сначала сегментация осуществляется в китайском тексте, затем в переводном русском тексте выделяются соответствующие элементы, называемые русскими синтаксическими аналогами (РСА) китайских клауз. Во второй части представлены принципы анализа и классификации РСА. В изученном материале разметки представлены 9 типов РСА китайских клауз: простые предложения, сложные предложения, группы предложений, фрагменты простых предложений, предложений с однородными сказуемыми, сложных предложений, части бессоюзных сложных предложений, сложносочиненных предложений и сложноподчиненных предложений. В третьей части сделаны выводы о том, что в большинстве случаев китайской клаузе может соответствовать монопредикативная конструкция в русском переводе. В других ситуациях китайские клаузы переводятся непредикативными словосочетаниями, полупредикативными причастными и деепричастными оборотами, полипредикативными сложными предложениями и группами предложений.

К л ю ч е в ы е с л о в а: клауза; параллельный корпус; дискурсивный корпус; дискурсивная разметка; выравнивание на уровне клауз; китайский язык; русский язык.

Ц и т и р о в а н и е: Ян И. Китайско-русский параллельный дискурсивный корпус: выравнивание на уровне клаузы и статистический анализ // Изв. Урал. федер. ун-та. Сер. 2: Гуманитар. науки. 2018. Т. 20. № 2 (175). С. 60–73.

Поступила в редакцию 19.01.2018

Принята к печати 18.04.2018

Yang Yi

Ural Federal University
Yekaterinburg, Russia**CHINESE-RUSSIAN PARALLEL DISCOURSE CORPUS:
ALIGNMENT OF CLAUSES AND STATISTICAL ANALYSIS**

This paper considers the alignment and annotation of Chinese-Russian parallel texts at the level of clauses. For the purpose of the article, the author creates a Chinese-Russian Parallel Discourse Corpus. As its pilot material, the author chooses an official Chinese document, namely, the 2017 Report on the work of the government of the PRC and its Russian translation. The article consists of three parts, in which the author 1) provides a description of the principles of alignment; 2) determines the annotation principles of the units selected and creates a parallel corpus with alignment at the level of clauses; 3) draws a statistic of the types of Russian syntactic analogues of Chinese clauses (RSACCs), and interprets the information received. The first part considers the way in which texts are divided into clauses in the parallel corpus; the author starts with the segmentation of clauses in the Chinese text, and then singles out elements in the Russian translation that are considered RSACCs. In the second part, the author describes the principles of analysis and classification of the Russian syntactic analogues. Overall, there are 9 types of RSACCs: simple sentences, complex sentences, groups of sentences, fragments of simple sentences, fragments of sentences with homogeneous predicates, fragments of complex sentences, parts of asyndetic complex sentences, and parts of complex and compound sentences. In the third part, the author concludes that in the majority of cases, Chinese clauses correspond to single-predicate constructions in the Russian translation; in other cases, Chinese clauses can be translated by means of non-predicative phrases, semi-predicative phrases, and poly-predicative complex sentences and groups of sentences.

Key words: clause; parallel corpus; discourse corpus; syntactic annotation; clause alignment; Chinese language; Russian language.

Citation: Yang, Y. (2018). Kitajsko-russkij parallel'nyj diskursivnyj korpus: vyravnivanie na urovne klauzy i statisticheskij analiz [Chinese-Russian Parallel Discourse Corpus: Alignment of Clauses and Statistical Analysis]. *Izvestia. Ural Federal University Journal. Series 2: Humanities and Arts*, 20, 2 (175), 60–73.

Submitted on 19 January, 2018

Accepted on 18 April, 2018

1. Введение

В современном дискурсивном анализе клауза понимается как элементарная дискурсивная единица (ЭДЕ). При создании дискурсивных трибанков (англ. *discourse treebanks*) сегментация текста связана с выделением клауз — например, *RST Discourse Treebank* [Carlson, Marcu, Okurowski], *Penn Discourse Treebank* [Prasad, Dinesh, Lee et al.], *Chinese Discourse Corpus* [Li, Feng, Sun et al.] и мн. др. Исследование машинного перевода показывает, что в системах Хуазцань

и Google Translation коэффициент точности перевода по клаузам (80 % и 65 %) гораздо выше, чем по предложениям (35 % и 20 %) [Wang]. Итак, установление соответствия дискурсивных единиц на уровне клауз становится одной из важнейших проблем для создания параллельных корпусов и улучшения машинного перевода.

В существующих параллельных корпусах выравнивание обычно предполагается на уровне предложений или абзацев. Что касается грамматической разметки на уровне клауз, то ее в современных китайско-русских параллельных корпусах почти нет. Учитывая предыдущий опыт при создании китайско-английского параллельного корпуса [Feng, Yang, Li et al.], мы предлагаем создание параллельного корпуса китайского и русского языков, в котором выравнивание осуществляется на уровне клауз.

В данной статье изучаются клаузы исходного языка (в нашем случае — китайского) и их соответствующие синтаксические аналоги переводного языка (русского). Сначала на основе параллельных текстов устанавливаются общие принципы выравнивания (разд. 2); далее на основе опыта разметки параллельного корпуса рассматриваются общие принципы соответствия русских синтаксических аналогов (РСА) китайским клаузам (разд. 3); наконец, приводятся статистические данные по разным типам РСА с интерпретацией полученной информации (разд. 4). В качестве пилотного материала был выбран китайский официальный документ «Доклад о работе правительства КНР 2017» и его перевод на русский язык.

2. Принципы выравнивания текстов на уровне клаузы

Выравнивание китайских и русских клауз в нашей работе начинается с сегментации пар текстов и включает следующие этапы: (1) в исходном китайском тексте по определенным принципам осуществляется деление на клаузы; (2) в переводном русском тексте производится деление на русские синтаксические аналоги, соответствующие китайским клаузам.

Понятие *клауза* мало изучается в традиционном синтаксисе китайского и русского языков. Для современного русского языка формальное определение клауз дают М. И. Ананьева и М. В. Кобозева при разработке корпуса текстов на русском языке [Ананьева, Кобозева]. В данной статье нас более интересует понятие клаузы в китайском синтаксисе, потому что выделение производится на основе исходного языка. Деление исходного китайского текста на клаузы учитывает три параметра: 1) ЭДЕ (в данном случае клауза) обязательно состоит из одного сказуемого и не является грамматическим элементом другой клаузы; 2) в ЭДЕ содержится как минимум одно суждение; 3) между ЭДЕ обычно ставится знак препинания (запятая, точка с запятой, точка и т. д.) [Мухин, Ян; Li, Feng, Zhou].

Русский текст делится в соответствии с сегментационными единицами китайских клауз, в результате получают русские синтаксические аналоги

китайских клауз (РСА). В примере (1) выделяемые единицы С1–С3 являются китайскими клаузами. Единицами R1–R3 являются РСА китайских клауз С1–С3. По форме R1 является целым предложением, R2 является частью сложноподчиненного предложения, а R3 — обособленный член сложноподчиненного предложения.

(1) ^{С1} 政府的一切工作都是为了人民,	^{R1} Вся деятельность правительства — во имя народа и для народа.	/
^{С2} 要践行以人民为中心的发展思想,	^{R2} Мы следуем концепции развития, в которой народ — центр внимания,	//
^{С3} 把握好我国处于社会主义初级阶段的基本国情。	^{R3} учитывая основные реалии Китая — нахождение на начальной стадии социализма.	

По этому примеру видно, что РСА могут быть различными синтаксическими единицами. Итак, перед нами стоит вопрос: каков перечень синтаксических единиц русского языка, которые могут выступать как функциональные аналоги китайских клауз? Для его решения важно проанализировать структурные особенности РСА и их грамматическую специфику.

Всего в ходе работы по сегментации получено 1 210 пар китайских клауз и русских синтаксических аналогов.

3. Общие принципы анализа русских синтаксических аналогов китайских клауз

В начале работы необходимо установить общие принципы анализа и классификации этих единиц.

Во-первых, анализ русских синтаксических аналогов обусловлен их структурной соотнесенностью с целым предложением, в конце которого ставятся завершающие знаки препинания. Например, в (1) РСА R2 и R3 определяются как части сложноподчиненного предложения.

Во-вторых, дальнейший анализ русских синтаксических аналогов обусловлен их частной грамматической спецификой. Подклассы выделяются на основе противопоставления частных моделей и, соответственно, грамматических значений.

Таким образом, анализ РСА учитывает как его общее структурное соотношение с целым предложением, так и его частную грамматическую специфику. Далее в разд. 4 рассмотрим классификацию РСА и проведем интерпретацию статистических данных.

4. Классификация РСА китайских клауз и статистика

4.1. Предложения, конструктивные части сложных предложений и фрагменты предложений

РСА по объему могут быть самыми разными: от фрагментов предложений до групп предложений. Под предложениями здесь мы понимаем предикативные

единицы, в конце которых ставятся завершающие знаки препинания. В состав предложений входят простые предложения, сложные предложения и группы предложений. Конструктивные части предложений являются составными частями сложных предложений, фрагменты являются любыми фрагментами, выделяемыми из предложений. На рис. 1 представлено количественное распределение типов РСА — предложений, частей сложных предложений и фрагментов предложений. Из рис. 1 видно, что количество фрагментов предложений (871) значительно превосходит количество предложений (339): они составляют 71,98 % всех русских синтаксических аналогов. Это свидетельствует о том, что при переводе с китайского на русский клаузам чаще соответствуют фрагменты предложений, а не целые предложения.

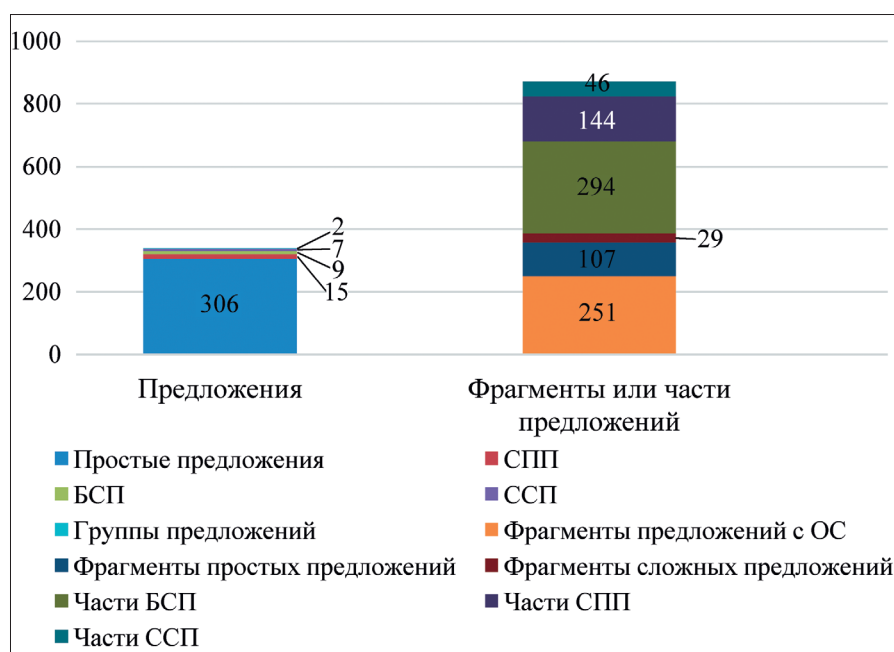


Рис. 1. Предложения, части сложных предложений и фрагменты предложений
 Pic. 1. Sentences, parts of complex sentences, fragments of sentences

4.2. Простые предложения

Простые предложения бывают двусоставными и односоставными. В разметке корпуса отражена классификация двусоставных предложений, основанная на структурном типе сказуемого. Соответственно, различаются три основных типа РСА: простые предложения с простым глагольным сказуемым (2), простые предложения с составным именным сказуемым (3) и простые предложения с составным глагольным сказуемым (4).

(2) 居民消费价格上涨2%。

Потребительские цены выросли на 2 %.

(3) 今年要基本完成公司制改革。

*В текущем году **будет** в основном **завершена** реорганизация госпредприятий в соответствии с Законом о компаниях.*

(4) —人民生活继续改善。

— *Продолжала **улучшаться** жизнь народа.*

С учетом способа выражения главного члена в односоставных предложениях выделяются глагольные и именные конструкции. В привлеченных текстах встретились инфинитивные (5), безличные (6), определенно-личные (7) и номинативные (8) предложения.

(5) 积极构建新型政商关系。

Форсировать формирование нового типа отношений между правительством и бизнесом.

(6) 必须持之以恒抓好安全生产。

Необходимо последовательно и настойчиво вести работу по обеспечению производственной безопасности.

(7) 向香港特别行政区同胞、澳门特别行政区同胞、台湾同胞和海外侨胞，表示诚挚感谢！

Выражаю искреннюю благодарность соотечественникам из СОАР и АОАР, тайваньским и зарубежным соотечественникам!

(8) (二) 深化重要领域和关键环节改革。

2. Углубление реформ в наиболее важных сферах и в ключевых звеньях.

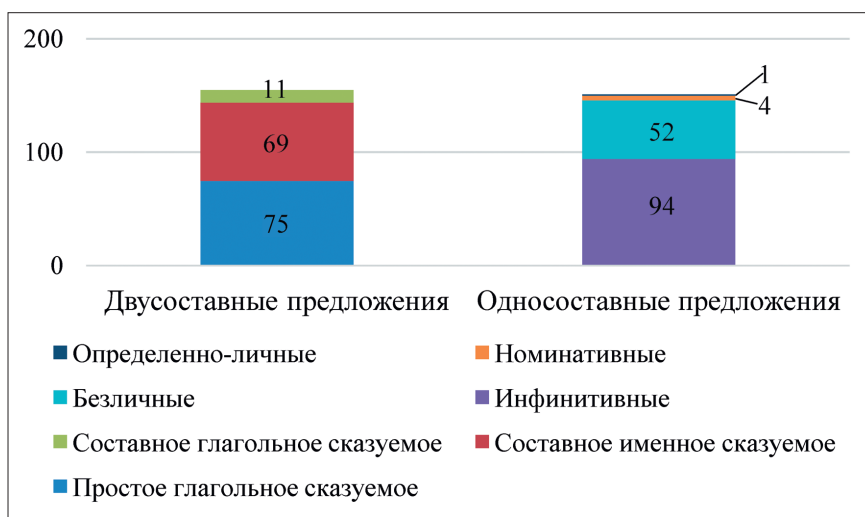


Рис. 2. Двусоставные и односоставные предложения

Рис. 2. Two-member and one-member sentences

Из данных на рис. 2 видно, что количество двусоставных (155; 12,81 %) и односоставных (151; 12,48 %) предложений в данной работе почти одинаково. Вместе они составляют 25,29 % всех РСА. Это подразумевает, что двусоставные и односоставные предложения одинаково важны при переводе китайских клаузов.

Во всех двусоставных предложениях простое глагольное сказуемое и составное именное сказуемое используются гораздо чаще, чем составное глагольное сказуемое. Среди односоставных предложений больше всего представлены инфинитивные конструкции, что является особенностью перевода официальных докладов с китайского на русский.

4.3. Фрагменты предложений

В рамки фрагментов предложений РСА могут входить главные члены предложения, второстепенные (в том числе обособленные члены). Например, в (9) R2 представляет собой главный член фрагмента простого предложения, а R1 — деепричастный оборот фрагмента простого предложения. Здесь следует отметить, что главные члены предложения в РСА китайских клауз могут быть переведены группами подлежащего и сказуемого отдельно. Например, в (10) аналоги R1 и R2 являются группами подлежащих, R3 — группой сказуемого.

(9) ^{C1} 一是继续创新和加强宏观调控, ^{C2} 经济运行保持在合理区间。

^{R1} Во-первых, продолжая внедрять инновации и усиливая макрорегулирование, ^{R2} мы сохранили функционирование экономики в рациональном диапазоне.

(10) ^{C1} 完善促进外贸发展措施, // ^{C2} 新设12个跨境电子商务综合试验区, / ^{C3} 进出口逐步回稳。

^{R1} Совершенствование мер по стимулированию внешней торговли // ^{R2} и создание 12 новых комплексных экспериментальных зон развития трансграничной электронной коммерции / ^{R3} привели к постепенной стабилизации импорта и экспорта.

Следует отметить, что фрагменты предложений могут быть результатом членения не только простых предложений, но и сложных. Например, в (11) аналоги R1–R3 определяются как фрагменты бессоюзных сложных предложений, R1 представляет собой группу сказуемого первой части предложения, R2 и R3 — второстепенные члены.

(11) ^{C1} 现在, 我代表国务院, // ^{C2} 向大会报告政府工作, /// ^{C3} 请予审议, / ^{C4} 并请全国政协各位委员提出意见。

^{R1} Разрешите от имени Госсовета // ^{R2} представить вам /// ^{R3} на рассмотрение доклад о работе правительства, / ^{R4} прошу членов Всекитайского комитета Народного политического консультативного совета высказать по докладу свои соображения.

РСА — фрагменты предложений также встречаются в предложениях с однородными сказуемыми, которые в традиционной лингвистике являются периферийными. В нашем корпусе предложения с однородными сказуемыми выделены в особую категорию. Например, в (12) русские синтаксические аналоги китайских клауз R1, R2 и R4 являются фрагментами предложения с однородными сказуемыми (*Необходимо претворять..., управлять..., выявлять...*), которые определяются как фрагменты безличного предложения.

(12) ^{C1} 我们要全面贯彻党的宗教工作基本方针, / ^{C2} 依法管理宗教事务, / ^{C3} 促进宗教关系和谐, // ^{C4} 发挥宗教界人士和信教群众在促进经济社会发展中的积极作用。

^{R1} Необходимо всесторонне претворять в жизнь основной курс партии в области религии, / ^{R2} управлять религиозными делами на основе закона, / ^{R3} способствуя установлению гармоничных межконфессиональных отношений, // ^{R4} выявлять позитивную роль религиозных деятелей и верующих в социально-экономическом развитии.

Из рис. 3 видно, что РСА-фрагменты чаще всего отмечаются в предложениях с однородными сказуемыми, что составляет 20,74 % всех РСА; в простых предложениях и в сложных предложениях — гораздо реже: 8,84 % и 2,40 %. Это говорит о том, что китайские группы сказуемых чаще всего переводятся в русском однородными сказуемыми. Между тем одна китайская клауза обязательно включает одну и только одну группу сказуемого.

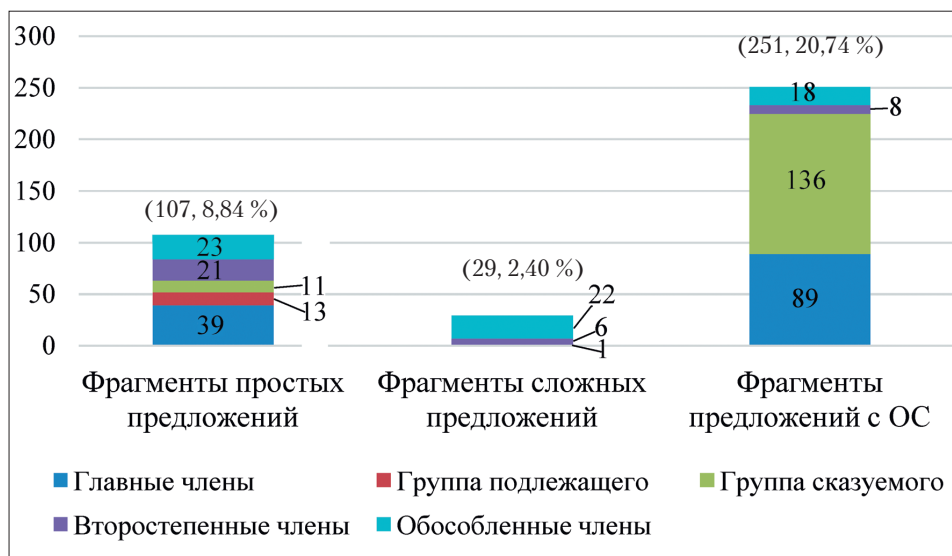


Рис. 3. Фрагменты предложений
Pic. 3. Fragments of sentences

4.4. Сложные предложения и группы предложений

РСА китайских клауз, которые переводятся как сложные предложения, в корпусе по традиции делятся на три группы: сложносочиненные (13), сложноподчиненные (14) и бессоюзные предложения (15).

(13) 贵在行动、成在坚持。

Действия очень важны, но успех зависит от наших неустанных усилий.

(14) 决不允许他们的辛勤付出得不到应有回报。

Ни в коем случае нельзя допускать, чтобы их нелегкий труд оставался неоплаченным.

(15) 加强法治宣传教育和法律服务。

Активизировать распространение правовых знаний и усиливать правовое воспитание, интенсифицировать оказание юридических услуг.

Кроме того, РСА китайских клауз также могут являться группами предложений (16).

(16) 坚持宏观政策要稳、产业政策要准、微观政策要活、改革政策要实、社会政策要托底的政策思路...

Необходимо продолжать обеспечивать устойчивость макроустановок, четкость производственной политики, гибкость микроустановок, практичность реформаторских мер и ориентированность социальной политики на закрепление базовых социальных гарантий. В этом заключается основная направленность наших политических установок.

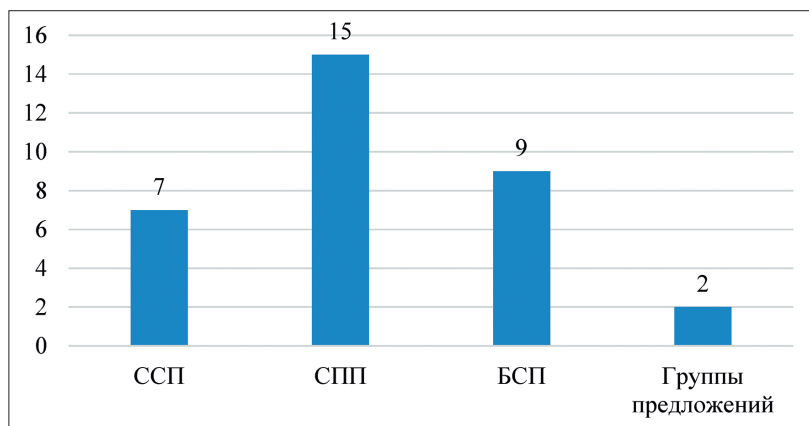


Рис. 4. Типы сложных предложений
Pic. 4. Types of complex sentences

В корпусе выделены 33 РСА, представляющих собой сложные предложения, что составляет 2,73 % всех РСА. На рис. 4 представлено количественное распределение сложных предложений. Таким образом, китайские клаузы могут быть переведены не только простыми предложениями русского языка, но и сложными, и даже группами предложений.

4.5. Конструктивные части сложных предложений

РСА китайских клауз могут быть формальными и смысловыми организациями простого предложения в составе сложных, названными в нашем исследовании конструктивными частями сложных предложений. В разметке корпуса классификация конструктивных частей сложных предложений, как и классификация отдельных простых предложений, основывается на структурном типе сказуемого. В итоге с конструктивными частями сложных предложений могут быть связаны следующие варианты разметки:

- двусоставное предложение:
 - (1) с простым глагольным сказуемым,
 - (2) с составным именным сказуемым,
 - (3) с составным глагольным сказуемым;

- односоставное предложение:

(4) инфинитивное,

(5) безличное,

(6) определенно-личное.

Например, в (17) R1 и R2 являются частями бессоюзного сложного предложения. Структурным типом сказуемого в R1 является простое глагольное сказуемое, а в R2 — составное именное сказуемое.

(17) ^{c1} 推进水污染防治, ^{c2} 出台土壤污染防治行动计划。

^{R1} Активизировалась работа по предотвращению и устранению загрязнения воды, ^{R2} опубликована программа действий, направленных на предотвращение и ликвидацию загрязнения почвы.

Из рис. 5 видно, что эти предложения могут быть двусоставными (250; 20,66 %) и односоставными (224; 19,34 %). В составе сложных предложений встречаются и двусоставные, и односоставные простые. Среди них при переводе чаще всего используются инфинитивные конструкции (180; 14,88 %) и предложения с простым глагольным сказуемым (152; 12,56 %).

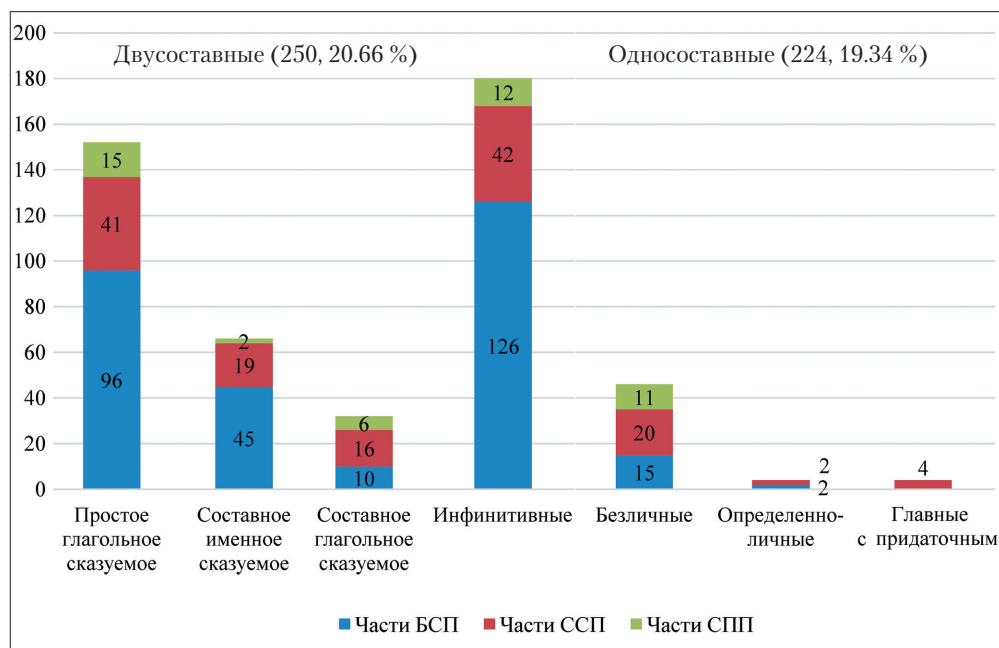


Рис. 5. Конструктивные части сложных предложений

Pic. 5. Construction parts of complex sentences

4.6. Предикативные, непредикативные, полупредикативные и полипредикативные конструкции среди РСА

В нашем корпусе среди РСА встречаются непредикативные, предикативные, полупредикативные и полипредикативные конструкции.

Предикативные единицы включают в себя обычные типы сказуемых: простое глагольное сказуемое, составное именное сказуемое и составное глагольное сказуемое. Кроме того, в корпусе еще различаются безличные, инфинитивные и определенно-личные главные члены, что позволяет отмечать конструктивные части сложных предложений, которые формально сходны с односоставными предложениями.

Непредикативные единицы являются словосочетаниями и предложно-падежными конструкциями, в которых не выражаются модальность, время и лицо.

Полупредикативные единицы являются обычно обособленными членами предложения: причастными и деепричастными оборотами.

Полипредикативные единицы проявляются в сложных предложениях и группах предложений.

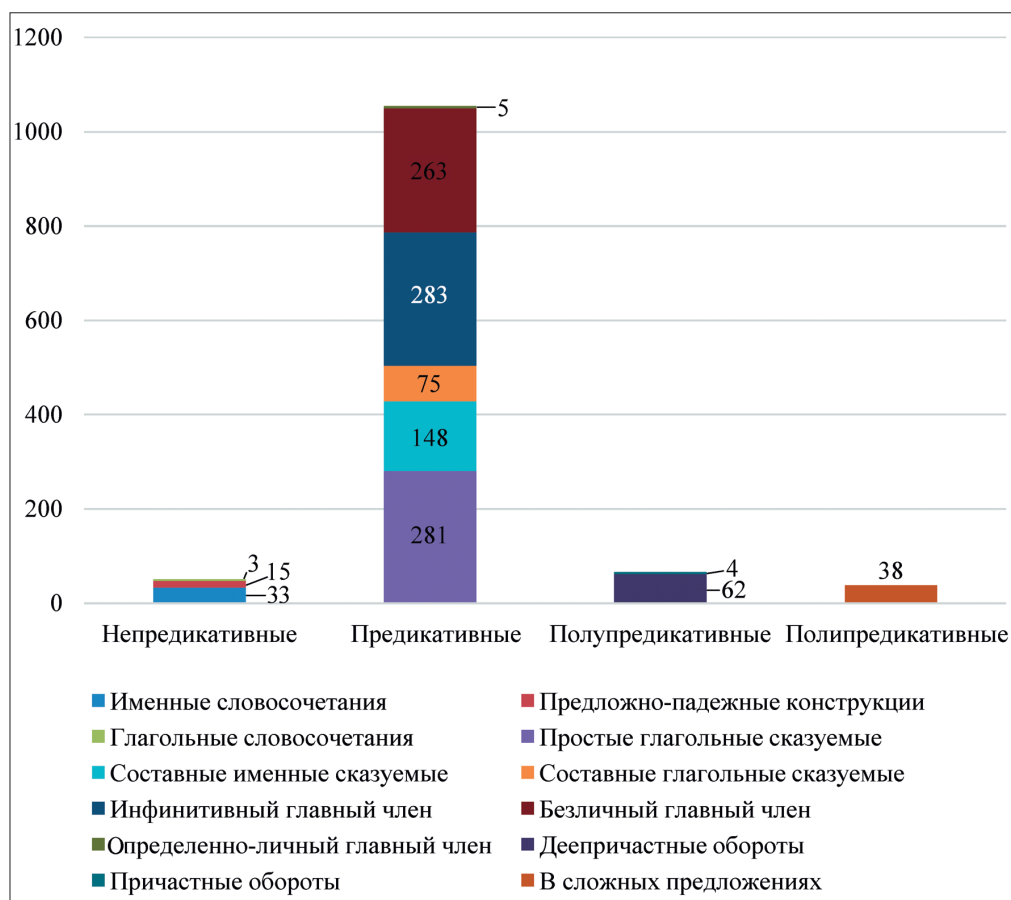


Рис. 6. Непредикативные, предикативные, полупредикативные и полипредикативные русские синтаксические аналоги китайских клаузов
 Рис. 6. Non-predicative, predicative, semi-predicative and poly-predicative Russian syntactic analogues of Chinese clauses

Как видно из рис. 6, предикативных единиц гораздо больше других (1043 случая): они составляют 86,2 % всех РСА. Это означает, что китайские клаузы обычно переводятся предикативными единицами русского языка. Поскольку китайская клауза обязательно состоит из одной группы сказуемого, то такой группе в большинстве случаев может соответствовать одна предикативная конструкция в русском переводе. В остальных случаях китайские глагольные группы переводятся непредикативными словосочетаниями (51; 4,21 %), полупредикативными причастными и деепричастными оборотами (66; 5,45 %), полипредикативными сложными предложениями и группами предложений (38; 3,14 %).

5. Выводы

1. В работе представлены принципы выравнивания китайских и русских текстов на уровне клауз. Следует отметить, что в данном случае выравнивание осуществляется только в направлении «китайский → русский», с учетом того, что китайский и русский языки отличаются друг от друга и по форме, и по структуре, поэтому разметка, одновременно осуществляемая в противоположных направлениях «китайский → русский» и «русский → китайский», не может быть адекватной.

2. Выявление и структурное описание грамматических особенностей русских синтаксических аналогов китайских клауз имеет важное значение в плане отбора подходящих синтаксических единиц при переводе. Однако вопрос о полном представлении грамматических особенностей китайских клауз пока остается открытым. В дальнейшем необходимо пополнить обсуждаемый корпус грамматической разметкой китайских клауз.

3. Данная работа строится на общетеоретической основе создания китайско-русского параллельного корпуса с дискурсивно-структурной разметкой. Выравнивание по клаузам является составной частью разметки на дискурсивном уровне, отражающей особенности дискурсивной структуры оригинального и переводного текстов. Цель создания такого корпуса соотносится с задачами машинного перевода, сопоставительного языкознания и методики преподавания иностранных языков.

Исследования

Ананьева М. И., Кобозева М. В. Разработка корпуса текстов на русском языке с разметкой на основе теории риторических структур [Электронный ресурс] // Компьютерная лингвистика и интеллектуальные технологии : по материалам ежегодной Международной конференции «Диалог» (Москва, 1–4 июля 2016 г.). М. : Изд-во РГГУ, 2016. Вып. 15 (22). URL: <http://www.dialog-21.ru/media/3460/ananyeva.pdf> (дата обращения: 29.11.2017).

Мухин М. Ю., Ян И. Проект создания китайско-русского параллельного корпуса официально-деловых текстов с дискурсивно-структурной разметкой // Вестн. ЮУрГУ. Сер. : Лингвистика. 2016. Т. 13. № 4. С. 23–31.

Carlson L., Marcu D., Okurowski M. E. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory [Electronic resource] // Current Directions in Discourse and Dialogue / eds. J. Van Kuppevelt, R. W. Smith. Kluwer, 2003. URL: <http://www.aclweb.org/anthology/W01-1605> (accessed: 28.11.2017).

Feng W., Yang Y., Li Y. et al. Study on the English Corresponding Unit of Chinese Clause // Natural Language Understanding and Intelligent Applications, Kunming, China, December 2–6, 2016, Proceedings / eds. C.-Y. Lin, N. Xue, D. Zhao, X. Huang, Y. Feng. Springer, 2016. P. 129–140.

Li Y., Feng W., Sun J. et al. Building a Chinese Discourse Corpus with Connective-driven Dependency Tree Structure [Electronic resource] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing / [ed. A. Moschitti]. URL: <http://www.aclweb.org/anthology/D14-1224> (accessed: 28.11.2017)

Li Y., Feng W., Zhou G. Research of Chinese Clause Identification Based on Comma // Acta Scientiarum Naturalium Universitatis Pekinensis. 2013. 49 (1). P. 7–14.

Prasad R., Dinesh N., Lee A. et al. The Penn Discourse Treebank 2.0 [Electronic resource] // Proceedings of the Sixth International Language Resources and Evaluation (LREC'08), Marrakech, Morocco, 28–30 May 2008. Marrakech: ELRA, 2008. URL: <https://www.seas.upenn.edu/~pdtb/papers/pdtb-lrec08.pdf> (accessed: 02.11.2017).

Wang J. Computer-oriented Chinese Translation Studies of English Clauses. Beijing: Beijing Language and Culture Univ. Press, 2009.

References

Ananyeva, M. I., & Kobozeva, M. V. (2016). Razrabotka korpusa tekstov na russkom iazyke s razmetkoi na osnove teorii ritoricheskikh struktur [Development of the Corpus of the Russian Language with Annotations Based on the Theory of Rhetorical Structures]. In *Komp'uternaia lingvistika i intellektual'nye tekhnologii: po materialam ezhegodnoi Mezhdunarodnoi konferentsii "Dialog" (Moskva, 1–4 iulia 2016 g.)* [Computer Linguistics and Intellectual Technologies] (Iss. 15 (22)). Moscow: RGGU Press. Retrieved from <http://www.dialog-21.ru/media/3460/ananyeva.pdf>. (In Russian)

Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In J. Van Kuppevelt, & R. W. Smith (Eds.), *Current Directions in Discourse and Dialogue*. Kluwer. Retrieved from <http://www.aclweb.org/anthology/W01-1605>.

Feng, W., Yang, Y., Li, Y., Li, X., & Ren, H. (2016). Study on the English Corresponding Unit of Chinese Clause. In C.-Y. Lin, N. Xue, D. Zhao, X. Huang, Y. Feng (Eds.), *Natural Language Understanding and Intelligent Applications* (pp. 129–140). Springer.

Li, Y., Feng, W., Sun, J., Kong, F., & Zhou, G. (2014). Building Chinese Discourse Corpus with Connective-driven Dependency Tree Structure. In A. Moschitti (Ed.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Retrieved from <http://www.aclweb.org/anthology/D14-1224>.

Li Y., Feng W., & Zhou G. (2013). Research of Chinese Clause Identification Based on Comma. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 49 (1), 7–14. (In Chinese)

Mukhin, M. Yu., & Yang, Y. (2016). Proekt sozdaniia kitaisko-russkogo parallel'nogo korpusa ofitsial'no-delovykh tekstov s diskursivno-strukturnoi razmetkoi [Building a Chinese-Russian Parallel Discourse Structure Corpus of Official Texts]. *Vestnik IuUrGU. Seriya: Lingvistika*, 13(4), 23–31. (In Russian)

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse Treebank 2.0. *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08), Marrakech, Morocco, 28–30 May 2008*. Marrakech: ELRA. Retrieved from <https://www.seas.upenn.edu/~pdtb/papers/pdtb-lrec08.pdf>.

Wang, J. (2009). *Computer-oriented Chinese Translation Studies of English Clauses*. Beijing: Beijing Language and Culture University Press. (In Chinese)

Ян И

аспирант кафедры фундаментальной
и прикладной лингвистики
и текстоведения;
младший научный сотрудник лаборатории
компьютерной лексикографии;
ассистент кафедры востоковедения
Уральский федеральный университет
620000, Екатеринбург, пр. Ленина, 51
E-mail: xwyang@mail.ru

Yang, Yi

PhD Student
Chair of Fundamental and Applied Linguistics
and Textual Studies;
Junior Researcher
Laboratory for Computational Lexicography;
Teaching Assistant
Department of Oriental Studies
Ural Federal University
51, Lenin Ave., 620000 Yekaterinburg, Russia
Email: xwyang@mail.ru
ORCID: 0000-0002-8715-1578
Scopus ID: 57192308753